

## CHAPITRE 20

# Evaluating Dutch Legal Research Quality Assessment

FABIAN AMTENBRINK

*Professor, Erasmus School of Law, Erasmus University Rotterdam  
and Visiting Professor, College of Europe, Bruges*

ALEX GEERT CASTERMANS

*Professor, Leiden Law School, University of Leiden and Director and  
fellow, Meijers Research Institute and Graduate School, Leiden\**

### I. Introduction

The aim of this contribution is to provide a critical overview of the current practice of legal research assessment in Dutch law schools and, thereafter, to discuss strategies towards the development of a sustainable research assessment system for the legal discipline.

In the first part the reader is introduced to the existing practice of research assessment at Dutch law schools. On the one hand, this practice is based on generally applicable standards defined by Dutch scientific and academic bodies, such as the Royal Netherlands Academy of Arts and Sciences (KNAW), the Netherlands Organisation for Scientific Research (NWO) and the Dutch Association of Universities (VSNU). On the other hand, each University and law faculty to a

\* The authors thank Jelle Oud LL.M for preparing several paragraphs of this paper. The views expressed in this contribution do not necessarily reflect those of the institutions with which the two authors are affiliated.

certain extent defines its own qualitative and quantitative standards as part of its research quality assurance system. This results in different – and arguably only partially congruent – practices. What is argued in this first part of the paper is that in the current situation a coherent research assessment system is missing in Dutch legal academia, as namely universally accepted methods to measure quality are (largely) absent, as is a ‘toolbox’ to actually measure impact and valorisation effects of legal research. These lacunae – arguably also to be found outside the Netherlands – are increasingly problematic, as government policy makers, (independent) funding bodies, as well as University managers ever more turn to measurable performance factors to decide on the allocation of scarce resources. This poses a major threat for legal research, calling for a rethinking of legal research assessment.

The second part of the paper then turns to possible strategies for the establishment of a sustainable research assessment system for the legal discipline. In this context two possible directions are discussed, namely the introduction of a system of quality indicators based on peer assessment and the use of bibliometric data to measure scientific output. As may be expected, both systems do not offer ready-made solutions and inter alia require a rethinking of the traditional channels of research distribution.

## **II. Research assessment in Dutch legal academia – an introduction**

### *A. Overview of the law school landscape in the Netherlands*

In order to provide some context for the discussion of the Dutch research assessment system, it is useful to provide a very brief overview of the general structure of Dutch law schools and the way they organise research.

#### *1. General structure of law schools*

In the nine Dutch law schools, legal education is organised according to the bachelor-master-system. A theoretical instruction of three years is leading to a bachelor’s degree (LL.B.), followed by a one or two year master’s programme degree (LL.M.). The members of the academic staff – PhD’s, assistant and associate professors, as well as the full professors – usually are involved in both teaching and research.

Nearly all law schools have special programmes or even research masters for talented students. These programmes are aimed at providing students with the necessary (methodological) skills to conduct further research after their graduation as Master of Laws, namely with the aim of obtaining a doctorate in law. PhD research positions usually are granted for three or four years. Specific training programmes for young researchers are often organised in the context of a Graduate School as part of the law school.

To a large extent, law schools are financed by the state, based both on the intake of students and the number of graduates. Research is partly financed by a surcharge on the fee for educational tasks (72 % of the research costs), and partly by external funding of individual researchers or research groups by national or European public funding authorities or third-party funded projects<sup>1</sup>.

## *2. Research programming as standard*

While previously legal research was mainly organised on an individual basis, from the end of the last century the programming of research has become common. This development has been considered necessary and at the time even inevitable, as law is more and more conceived 'in context' and entailing a multi- or interdisciplinary approach. Legal researchers no longer confine themselves to one traditional field of law, such as private or public law, but rather embrace other fields of law as well, both nationally and internationally. In this changing landscape, the programming of legal research and the focus on research groups is considered more efficient, effective and to produce a more profound output than could be achieved by individuals. Highlighting the scale of this development, at the time of the nationwide external research evaluation in 2009, the nine law schools ran some sixty research programmes, ranging from small programmes with very specific research goals to large programmes with broadly defined areas of research and without specific goals<sup>2</sup>.

1. For details, see Evaluatiecommissie Rechtswetenschappelijk onderzoek (Commissie Koers), *Kwaliteit en diversiteit* (Report), Amsterdam 2009, p. 36.

2. Evaluatiecommissie Rechtswetenschappelijke onderzoek (Commissie Koers), *Kwaliteit en diversiteit* (Report), Amsterdam 2009, pp. 42 and 43.

*B. The Dutch (legal) research assessment system*

While there are considerable differences between the nine Dutch law schools not only with regard to the substantive focus of research but also to the organisational framework, a generally applicable research system exists to some extent.

*1. VSNU, KNAW, NWO, Standard Evaluation Protocol (SEP) 2009-2015*

For reasons of accountability (the Ministry of Education, Culture and Science is the biggest financer of academic research in the Netherlands) and so to provide insights into their research activities, since 1992 Dutch universities are by law required to ensure a regular assessment of the quality of the output of their institutions<sup>3</sup>. To this end, general protocols for quality control of academic research have been developed by the KNAW (advisory body of scientists to the government), the NWO (state controlled and sponsored organisation for the allocation of research resources) and the VSNU (advocate of Dutch universities). At the time of writing of this contribution, the Standard Evaluation Protocol (SEP) 2009-2015 is used to assess all academic research at Dutch universities, and thus has not been specifically established for the assessment of legal scholarship.

According to the SEP, the assessment of research consists of a self-evaluation, which is subject to an external review by a national evaluation committee, including a site visit once every six years, and an internal midterm review in between two external reviews. The external evaluation of scientific research is to be applied at two levels: the research institute as a whole and its research programmes. The assessment entails four main criteria, i.e. quality, productivity, relevance, and vitality & feasibility<sup>4</sup>, whereby the self-evaluation report has to reflect on these issues. The report must also contain an analysis of the institute's strengths and weaknesses and perspectives on the future, and a full set of quantitative information concerning the input and output, including key publications in order for the evaluation committee to determine the quality and scientific relevance of an institute or research programme.

3. See the *Wet Hoger onderwijs en Wetenschappelijk onderzoek*, art. 1.18, art. 2.5. par. 2.

4. Standard Evaluation Protocol (SEP) 2009-2015, pp. 9-11.

## EVALUATING DUTCH LEGAL RESEARCH QUALITY ASSESSMENT

The SEP is characterized by a quantitative approach (see the full set of quantitative information concerning the input and output) and a strong focus on the beta sciences. This is apparent for instance from the method of quantification of quality, which places a strong emphasis on refereed articles in highly ranked journals and little emphasis on other categories of research output and professional publications (table 1), and a focus on ‘programmed research’.

Table 1: Categories of research output at institutional and programme level in the SEP

Entire Institute or research programme
Refereed articles
Non-refereed articles (1)
Books
Book chapters
PhD-theses
Conference papers
Professional publications (2)
Publications aimed at the general public
Other research output <specify> (3)

Note 1: Articles in journals that are non- refereed, yet deemed important for the field

Note 2: Publications aimed at professionals in the public and private sector (professionele publicaties), including patents and annotations (e.g. law).

Note 3: Other types of research output, such as abstracts, editorships, inaugural lectures, designs and prototypes (e.g. engineering) and media appearances.

From the outset it has been clear that the SEP needed to be geared towards each scientific discipline; in anticipation of the SEP 2009-2015, the KNAW published a rapport in which was stated that two issues were to be improved in the SEP 2009-2015, one of which was that assessment procedures should do more justice to ‘the variety of disciplines and developments in research and should reflect the differences in research practices’<sup>5</sup>.

According to the KNAW, great differences exist between scientific disciplines with regard to ‘publication and communication practices

5. Academy Committee for Quality Assurance, *Quality assurance in scientific research : From Sep to CEP*, Amsterdam, KNAW, 2008, p. 6.

and interaction with the actual practices within the community'<sup>6</sup>. Also, more specifically related to publication behaviour, the KNAW stated that 'with regard to the assessment of scientific publications, differences between scientific disciplines concerning publication conventions are hardly taken into account'<sup>7</sup>. Thus, the differences between scientific discipline with regard to publication traditions and outlets has been identified as a fundamental obstacle for a uniform assessment of the quality and impact of publications<sup>8</sup>.

During the preparation of this contribution, the successor to the SEP 2009-2015 has been announced: the SEP 2015-2021<sup>9</sup>. In some areas, the new SEP differs significantly from its predecessor. For instance, in the new SEP 'productivity' ceases to be an independent criterion, as the focus in terms of assessment criteria comes to rest on three basic pillars: research quality, relevance to society and viability. Also, comparative (national) assessments will no longer be compulsory. In addition, the new SEP demands a further elaboration of quality indicators for the societal relevance of research. However, despite these adjustments, the SEP 2015-2021 does not fill the aforementioned lacunae.

As the basic principles of the old and new SEP remain the same and considering that the SEP 2015-2021 has not yet been put into operation yet, this contribution hereafter focuses on the SEP 2009-2015<sup>10</sup>.

## 2. KNAW, *Quality and relevance in the humanities, Advisory Commission Quality indicators for the Humanities 2012*

With regard to the assessment of the humanities in the Netherlands, the Advisory Commission Quality Indicators for the Humanities observed in 2011 that although there was no reason to adopt a different approach to the concept of research quality to that adapted in other fields of science and scholarship, the humanities however

6. *Ibid.*, pp. 16-17.

7. DE JONGE AKADEMIE, *Kennis over publiceren. Publicatietradities in de wetenschap* (Advice), KNAW, Amsterdam 2012, p. 7.

8. *Ibid.*, pp. 8-9.

9. <https://www.know.nl/nl/actueel/publicaties/standard-evaluation-protocol-2015-2021> (all websites cited in this article have been accessed on 29 September 2014).

10. Thus, unless stated otherwise, hereafter the term « SEP » refers to the SEP 2009-2015.

‘do demand a fairly wide range of quality indicators that do justice to the diversity of products, target groups, and publishing cultures’ within this field<sup>11</sup>.

To this end, and because of the limited usefulness for the majority of the humanities of for instance indicators based on citation indexes<sup>12</sup>, a suitable method for evaluating humanities research, based on peer review was developed and tested in the Netherlands with a lot of positive response, ‘in particular because it recognizes the versatility of research in the humanities (through the creation of indicators for each criterion that can be used in a field-specific way, red.) while avoiding unnecessary complexities and remaining transparent’<sup>13</sup>.

### 3. KNAW, *Towards a framework for the quality assessment of social science research*, “Quality Indicators in the Social Sciences” Committee, March 2013

In 2013, the *Quality Indicators in the Social Sciences Committee* took a clear position on some basic principles concerning the assessment of quality of social science research. One important principle was that quality assessment ‘is based on a set of *generally* applicable assessment criteria in which *specific* quality indicators and their weights can be identified and defined’<sup>14</sup>, since a set of generally applicable quality indicators ‘will not make enough allowance for the diversity of social science research’<sup>15</sup>.

Another interesting principle states that ‘within this set of generally applicable assessment criteria, researchers and research groups have the right to identify quality indicators suitable for their own field, and the obligation to account for these indicators and how they are weighted’<sup>16</sup>.

11. COMMITTEE ON QUALITY INDICATORS IN THE HUMANITIES, *Quality indicators for research in the humanities* (Interim report), Amsterdam, KNAW 2011, p. 11.

12. ADVIESCOMMISSIE KWALITEITSINDICATOREN GEESTESWETENSCHAPPEN, *Kwaliteit en relevantie in de geesteswetenschappen* (Eind rapportage), Amsterdam, KNAW 2012, p. 13.

13. *Ibid.*, p. 11.

14. QUALITY INDICATORS IN THE SOCIAL SCIENCES COMMITTEE, *Towards a framework for the quality assessment of social science research* (Advisory report), Amsterdam, KNAW March 2013, p. 9.

15. *Ibid.*, p. 17.

16. *Ibid.*, p. 9.

4. *The Standard Evaluation Protocol for legal research 2005-2009*

Because, as has been mentioned above, the various scientific disciplines differ in approach, methodology, publication and debate culture (including bibliometrics) and finance<sup>17</sup>, a special Committee (*‘Oordelen over rechten’*) was requested to develop a coherent view on four issues that needed special attention with regard to the assessment of legal research:

- a workable classification of scientific and professional publications;
- the meaning and consequences of internationalization for legal research;
- the possibility and the desirability of a classification of legal journals;
- the possibility and the desirability of performance indicators of legal research<sup>18</sup>.

This resulted in the *Standard Evaluation Protocol for legal research 2005-2009*, a protocol specifically designed for the evaluation of legal research, which primarily sought to improve the comparability between law schools with regard to the set of quantitative information<sup>19</sup>.

The SEP for legal research solved some of the ‘burning’ issues at stake. With regard to the method of quantification of quality for instance, specific characteristics of the publication culture in legal research were taken into account (see table 2).

Table 2: Programme results in numbers in the SEP for legal research

Scientific publications	Professional publications	Popular publications
Monographs	Monographs	All publications
PhD-theses	Books	
Books	Editorship	
Articles in externally refereed journals	Articles	
Articles in other journals	Annotations	
Editorship		
Conference paper		
Other		

17. Commissie Voorbereiding Onderzoeksbeoordeling Rechtsgeleerdheid, *Oordelen over rechten* (Report), VSNU 2005, p. 1.

18. *Ibid.*, p. v.

19. *Standaard Evaluatie Protocol Rechtsgeleerdheid 2005-2009*, p. 1.



Law schools were also provided with binding classification criteria for scientific and professional publications. A scientific publication has been defined as a ‘publication about a result of academic research, aimed at the forum of researchers’, which must comply with the requirements of originality, thoroughness and profundity<sup>20</sup>. The characteristic distinction with professional publications is that scientific publications ‘*increase* the body of knowledge’, whereas professional publications, which are defined as publications about a result of academic research, aim at an audience with professional interests and as such ‘disseminate and stimulate the application of *existing* knowledge’<sup>21</sup>. However, the SEP for legal research neither contains a classification of legal journals nor provides for concrete indicators for quality. Thus, to a large extent the SEP for legal research remained dependant on the general SEP<sup>22</sup>.

*5. Outcomes of the report of the Smits committee on quality indicators and ranking and the (interim) findings of the Du Perron committee (a follow-up to the Smits committee) on how to determine the quality of legal journals*

Consequently, on the basis of the rapport (‘*Oordelen over rechten*’), two committees were created. One committee<sup>23</sup>, which was given the task to develop quality indicators for legal research, argued that good legal research must be at least original, thorough and profound, that peer review is the most reliable assessment method and that quality indicators have to be developed to ‘objectify’ the process of peer review. The Smits Committee also made some provisions regarding the ranking of legal journals<sup>24</sup>. A follow-up committee was supposed to develop a ranking of Dutch legal journals<sup>25</sup>. However, the proposals in the interim report were not warmly welcomed and supported in the field. This might not come as a surprise since the chairman of

20. COMMISSIE VOORBEREIDING ONDERZOEKSBEOORDELING RECHTSGELEERDHEID, *Oordelen over rechten* (Report), VSNU 2005, pp. 29-30.

21. *Ibid.*, p. 32.

22. *Standaard Evaluatie Protocol Rechtsgeleerdheid 2005-2009*, p. 1.

23. Commissie Prestatie-indicatoren en ranking (“Commissie Smits”), VSNU.

24. J.B.M. VRANKEN and R.A.J. VAN GESTEL, « Het dilemma bij de beoordeling van rechtswetenschappelijke publicaties: Vertrouwen is goed, controle niet altijd beter », *Ars Aequi* 2010 (59) 3, p. 167-168.

25. Commissie Classificatie Juridische Tijdschriften (« Commissie Du Perron »), VSNU.

the committee stated that there are hardly any (truly) *scientific* legal journals in the Netherlands<sup>26</sup>. A final report has yet to appear.

Despite the efforts of both committees, their findings have not been shared nor widely discussed within the legal discipline<sup>27</sup>.

#### 6. *Observations on the assessment of Dutch legal research 2003-2009*

In 2009, the Executive Board of nine Dutch universities with law schools established a Committee to evaluate the quality of legal research in the years 2003-2009, in compliance with the SEP and the SEP for legal research. This subsequently resulted in the assessment of all research institutes and their programmes.

Needless to say, assessing the output of nine law schools for a period of seven years is a mug's game. Lacking a (objective) system of peer review and ranking of journals, it became a research project on its own. Although the members of the Committee were acquainted with a large number of publications, doubts were raised whether the Committee had actually achieved to assess a cross sample of all research. In its final report, the Committee observed that this external assessment was so time consuming – not only for the members of the Committee but, foremost, for the academic (support) staff at the Universities – that it was doubtful whether it could be organised in a similar manner in the future<sup>28</sup>. Interestingly these concerns were taken into account in the development of the 'new' SEP 2015-2021. Alternatively to a country-wide comparative research assessment, the protocol now also allows for stand-alone research assessments at the level of the individual research institution<sup>29</sup>.

Given the assessment criteria (in particular the criterion of productivity), the strong quantitative approach in SEP 2009-2015 and the absence of generally shared quality indicators within the legal research community, it might not come as a surprise that according

26. C.E. DU PERRON, "De kwalificatie van juridische tijdschriften", in G. VAN DIJCK, R.A.J. VAN GESTEL, I. GIESEN and A. HAMMERSTEIN (Eds.), *Cirkels: Een terugblik op een vooruitziende blik*, Deventer, Kluwer, 2013, p. 194.

27. J.B.M. VRANKEN and R.A.J. VAN GESTEL, "Het dilemma bij de beoordeling van rechtswetenschappelijke publicaties: Vertrouwen is goed, controle niet altijd beter", *Ars Aequi* 2010 (59), 3, p. 179.

28. EVALUATIECOMMISSIE RECHTSWETENSCHAPPELIJKE ONDERZOEK (Commissie Koers), *Kwaliteit en diversiteit* (Report), Amsterdam 2009, pp. 48-50.

29. Explanatory memorandum to the *Standard Evaluation Protocol 2015-2021*.

to the Committee an important weakness of Dutch legal research is the increasing emphasis in recent years on quantity instead of quality. This encourages strategic behaviour by researchers that may be tempted to submit three shorter articles instead of one large one. The Committee therefore questioned whether legal research benefits from such a focus on quantity<sup>30</sup>.

Indeed, without quality indicators, such as the journal rankings, citation indexes and double blind peer reviewed journals to be found in other disciplines, assessment of legal research is mainly based on the publication form, medium (journals, book etc.) and/or the language in which a publication is written. Yet, in principle these factors say little about the intrinsic quality of a publication<sup>31</sup>.

Apart from productivity and quality, societal relevance is considered to be an import criterion. Traditionally, legal research is strongly connected to legal practice. In its final report, the Committee argues that this connection is at danger. It notes a tension between legal research and legal practice, referring to the pressure on researchers to produce 'scientific publications at the cost of professional publications' and to the declining number of professional publications. It argues that a focus on the number of scientific publications as well as on internationalisation might come at the cost of publications that are fitted to legal practitioners<sup>32</sup>. Here, too, it must be admitted that there is no general understanding on what the valorisation of legal research means, what the importance of it is and/or should be and how it can be measured<sup>33</sup>.

Concluding, from the 2009 external research evaluation it has become clear that the publication strategies and corresponding quality assurance procedures (peer review, ranking, scientific impact), as well as the societal impact (valorisation) of legal scholarship are the most important issues that need to be addressed by the legal discipline<sup>34</sup>.

30. EVALUATIECOMMISSIE RECHTSWETENSCHAPPELIJKE ONDERZOEK (Commissie Koers), *Kwaliteit en diversiteit* (Report), Amsterdam 2009, pp. 54-59. As has been mentioned before, "productivity" ceases to be an independent criterion in the (new) SEP 2015-2021.

31. J.B.M. VRANKEN & R.A.J. VAN GESTEL, « Het dilemma bij de beoordeling van rechtswetenschappelijke publicaties : Vertrouwen is goed, controle niet altijd beter », *Ars Aequi* (59) 3, 2010, p. 176.

32. EVALUATIECOMMISSIE RECHTSWETENSCHAPPELIJKE ONDERZOEK (Commissie Koers), *Kwaliteit en diversiteit* (Report), Amsterdam 2009, pp. 39-40.

33. This issue is elaborated in the SEP 2009-2015.

34. C.J.J.M. STOLKER, « Een discipline in transitie. Rechtswetenschappelijk onderzoek na de Commissie Koers », *Recht en Methode in onderzoek en onderwijs* 2011, p. 38.

### C. *Intermezzo*

#### 1. *A case study: assessing an individual publication*

A scientific publication in the Netherlands is defined as a ‘publication about a result of academic research, aimed at the forum of researchers’, increasing the body of knowledge and complying with the requirements of originality, thoroughness and profundity. The last three – content based - criteria are widely accepted criteria for the assessment of publications through research assessment exercises<sup>35</sup>, yet the manner of interpretation of these criteria may differ<sup>36</sup>. Subsequently, a methodological justification for research questions and problem definition, proper citations and consistency are also desirable<sup>37</sup>.

To illustrate the difficulties of assessing a specific publication, these criteria are applied to the publication ‘Towards a European Contract Law through Social Dialogue’, by Alex Geert Castermans, published in the *European Review of Contract Law* 2/2011, pp. 360-367. This article is presently chosen, because in the context of a mere illustration of the debate, it is considered appropriate to question the work of one the authors of this contribution rather than the work of another colleague.

Abstract: Suppose the Draft Common Frame of Reference – a set of rules and principles in the field of private law, drawn up by a consortium of European academics in the field of private law – would be chosen as an optional instrument, will it make business easier? It will not, because of its various references to national law, both of a private and public law nature. It turns out a perilous undertaking to offer private parties a set of rules, by way of an optional instrument, because it is created quite apart from the national private-law and public-law context and it covers only part of the private-law legal relationship. In every single case such a set of rules must be embedded in the applicable national private and public law. An alternative

35. COMMISSIE VOORBEREIDING ONDERZOEKSBEOORDELING RECHTSGELEERDHEID, *Oordelen over rechten* (Report), VSNU 2005, pp. 29-31.

36. J.B.M. VRANKEN and R.A.J. VAN GESTEL, « Het dilemma bij de beoordeling van rechtswetenschappelijke publicaties : Vrouwen is goed, controle niet altijd beter », *Ars Aequi* (59) 3, 2010, pp. 168-171.

37. J.B.M. VRANKEN and R.A.J. VAN GESTEL, « Rechtswetenschappelijke artikelen. Naar criteria voor methodologische verantwoording ». *Nederlands Juristenblad*, 82(24), 2007, 1451.

is explored. A European Self-regulatory consultation system could provide a framework for drafting model contracts that are tailored to individual sectors. These models should offer terms and conditions that are considered acceptable throughout Europe, with alternatives for Member States having wider or narrower margins.

The article is written in English and published in an international journal. Yet, whether and to what extent the publishing of a paper in another language, mainly English, says something about the intrinsic quality of the output, is unclear in the Netherlands.

The article is supposed to add something to the body of knowledge, because it tests the presumption that a common European Private Law would make doing business easier from the perspective of a multilevel jurisdiction. Are therefore the basic requirements for legal search being met, for instance with regard to originality, depth and thoroughness? Is there a clear problem definition and justification for the use of all sources?

It is a written along the lines of a story, the beings and doings of a wine merchant, confronted with various rules and regulations concerning the quality of wine. It takes a few lines before a research question is to be discovered, taking as a starting point the hypothesis that with a European body of private law European consumers will feel as safe when doing business with a foreign company as they would feel with companies from their own country, and that Europe's small and medium-sized companies can offer their products and services to consumers in other countries without having to become experts in the national contract law systems of all other EU countries. This hypothesis is tested by assessing the legal position of the wine merchant, in particular to which extent it is depending on the proposed common European Private Law and to which extent it is depending on national law, either private or public national law.

Does this then amount to a thorough and profound legal analysis? Is it original and transboundary? Certainly, not all sources have been used. Many articles have been written on the desirability of a Common European Private Law. Some articles try to uncover its economic value. Other research assesses the level of consumer protection. They are not mentioned. The article aims at a systematic analysis of the Draft Common Sales Law, from a multilevel jurisdiction perspective. It is open to debate whether it is necessary to go into all the details of the subject.

The article has been published in an international peer reviewed journal, and therefore classified as an article in a journal; however, it's unclear how the editors define peer review (by editors or external peers?) and to what extent they pay attention to the basic requirements for legal research mentioned above.

Finally, does the evaluation of this article depend on its impact and valorisation? It is available in the Leiden University Repository as well as on the website of the European Economic and Social Council. It has been submitted in the consultation with regard to the green paper on policy options for progress towards a European Contract Law for consumers and businesses (COM(2010) 348) and it is referred to in some other publications<sup>38</sup>. Statistics of the Leiden University Repository learn that it has been viewed 207 times, downloaded 73 times.

Despite the disappointing statistics, the publication is booked as 'scientific'. It was one out of several publications that have been considered scientific, so, the author, working at Leiden Law School, passed the internal quality assessment. He became fellow at the E.M. Meijers Institute. His research programme, Coherent Private Law, has been assessed recently, with a positive result. But what does this all mean if one feels the need to compare the result with the work of fellow researchers and research groups?

## *2. A case study; the Leiden Law School Midterm Review 2009-2012*

Exemplary for the practical challenges involved in operationalising a standard protocol for the evaluation of the research of a law school is the Leiden Law School Midterm Review which took place in 2013, covering the period 2009-2012.

The Leiden Law School consists of five academic institutes and has five research programmes, which more or less overlap each other. Nearly all research within Leiden Law School takes place within these programmes and institutes, which in their turn are part of the faculty research programme 'Interaction between legal systems'. Together, the five research programme coordinators, a representative of the PhD-students and the Leiden Law School Director of Research

38. F.E. GERHARD DANNEMANN, S. VOGENAUER (Eds.), *The Common European Sales Law in Context*, Oxford University Press 2013.

form the Research Board, which is responsible for the organisation, development and progress of research. The Research Board is assisted in this task by the Meijers Research Institute and Graduate School<sup>39</sup>.

The first phase of the Midterm Review started with a day of common reflection ('toogdag') by the participants of every research programme. In addition, each research programme had to draw a self-assessment report conform the (requirements of the) SEP 2009-2015, on which a second day of reflection followed, this time with external reviewers (peers). On the basis of the self-assessment report and the 'visit', the external reviewers passed a judgement on the individual research programmes. In this phase, which took both the past, present and future of the research programme into account, the focus came to rest on the general content of the research programme and less on the results of individual members (see par. II.D.1.). For instance, main focus points in this phase were the research programme's research questions, its (new) perspectives, societal challenges, as well as internal coherence.

In the second phase of the Midterm, the Meijers Research Institute and Graduate School had to draw a self-assessment report, also in line with the (requirements of the) SEP 2009-2015. Again, a day of reflection with external reviewers (peers) followed. During this 'visit', the external reviewers among other things held interviews with the Research Board, the 'Meijers fellows' (see par. II.D.1.) and PhD students. The second phase also ended with a review-report by the external reviewers, based on the self-assessment report and the visit.

Despite the absence of clear definitions ('scientific quality', 'societal relevance' etc.) and generally accepted and fully elaborated assessment criteria, the research programmes and Meijers Research Institute and Graduate School received overall positive reactions and recommendations. In general, the Midterm Review led to fruitful discussions within the Research Board, the institutions, the research programmes and the Faculty Board on topics such as the organization of legal research (autonomy vs. centralization). In addition, it resulted in clearer choices on certain issues and led to changes of the research focus in some programmes and the 'strengthening' of others.

The second phase (and thus the whole assessment procedure) ended with a formal response by the Faculty Board. In this response,

39. However, the institute is also responsible for both the Master Talent and PhD Programmes.



the Board stated its own position on the assessment outcomes and expressed the consequences it attached to it. The assessment report and response by the Faculty Boards were then published.

Although the midterm constituted an inspiring exercise, without general consensus within the Dutch legal academic about what quality of legal research is and how it can be assessed, it did not and could not provide for a coherent research assessment system.

*D. Law school specific qualitative and quantitative output norms – the examples of Leiden and Rotterdam*

In the absence of a clearly defined and commonly expected standard for quality assurance, each law school has to define its own standards. The Law School of the University of Leiden and the Erasmus School of Law of the Erasmus University of Rotterdam may serve as examples in this regard.

*1. Research quality assurance systems at Leiden Law School*

With the establishment of the Meijers Research Institute and Graduate School in 1996, a quality assessment system for scientific research was developed. To fully understand this system it is important to make a distinction between the different groups of academic staff: post doctorate-researchers, PhD candidates and other researchers.

Once every three year, within the framework of fellow reviews, post-doctorate researchers are assessed by the Board of the Meijers Research Institute and Graduate School on the quality and extent of their scientific research. It is important to note that the fellow status is included as a prerequisite in the appointments of university lecturers, senior university lecturers and professors. The first series of reviews was held in 1998 and covered the period 1996-1997. The last general set of reviews took place in 2013 (2010-2012).

For potential fellows of the Institute, the Board applies norms derived from the SEP. Like the SEP, the quality assessment system for scientific research at Leiden Law School is characterised by a strong quantitative approach. A full-time staff member should produce seven scientific articles every three years, including one article in an international review. This corresponds with the standard that's being used by the VSNU to assess Dutch law schools. For the distinction between scientific and professional publications reference is



made to the SEP for legal research<sup>40</sup>. Classification is left to the individual researcher but has to be approved by the research programme Directors. To enable the assessment of output of individual researchers, every researcher has to submit his or her publications, classified according to a classification of research output based on both the general as the SEP for legal research, into LUCRIS, an electronic database which is also used to produce the set of quantitative information for the general assessment.

Apart from a specific output, the aspiring fellows will also be asked to produce a brief written account of their results (including the record of external financial support) and plans for the coming years. This serves to clarify the opportunities for mutual cooperation within the separate programmes and also between the programmes.

The performance of PhD candidates is assessed in the progress discussions with their supervisors and co-supervisors. The Leiden Law School Dean of PhD Studies monitors this process and also personally conducts regular discussions with the PhD candidates and their supervisors. Concerning other researchers such as the project staff, assessment takes place only within the framework of annual performance interviews at the level of one of the five Academic Institutes of Leiden Law School. Performance interviews are held with all members of academic staff, including professors.

Besides the assessment of individual staff members, a midterm review is used to assess the research programmes and the Graduate school (as has been mentioned before). The sole purpose of the midterm review, which takes place between two external reviews and is performed by external specialists per research programme, is to determine the current state of affairs regarding the research institute as a whole and its research programmes: what has been done with the comments of the last assessment committee, and what has to be done or what has to be improved in the light of the upcoming assessment procedure?

40. *Onderzoekprotocol van de Leiden Law School ten behoeve van de fellowronde over de periode 2013, 2014 en 2015*, pp. 2 and 3.

## 2. Research quality assurance system at Erasmus School of Law<sup>41</sup>

The research quality assurance at the Erasmus School of Law is *inter alia* geared towards ensuring:

- Minimum scientific qualification of staff with research time;
- Adequate scientific output of individual researchers in terms of quality and quantity;
- Adequate scientific output of individual researchers contributing to the research programmes and frontier research groups of the faculty;
- Efficient allocation of financial resources;
- Periodical impartial evaluation of the research performance of individuals and at the institutional level;
- Selection and training of PhD candidates.

In principle only academic staff with a PhD or equivalent qualification can qualify for research time. In order to be allotted research time and the corresponding financing, researchers have to obtain a research qualification. This research qualification is also a *conditio sine qua non* for participation in one of the faculty's research programmes and frontier groups.

The research qualification requires researchers to produce an average of three scientific publications per year in a consecutive period of three years. All publications have to be entered into an electronic database, which mirrors the classification of output in the before-mentioned SEP. As a generally accepted definition for refereed and non-refereed journals for the legal discipline is missing in the Netherlands, Erasmus School of Law has established its own refereed journals list, which is based on the list established by consensus for the purpose of the 2009 Dutch legal research assessment. This list is then supplemented with journals that meet the standard set out in the definition of the Association of Universities in the Netherlands (VSNU) for refereed publications<sup>42</sup>. In addition, the ESL journal list identifies the number of publications in journals and the related impact factors according to common indexes in Legal Sciences: the Foreign Legal Periodicals Index (IFLP)<sup>43</sup>; the Washington & Lee

41. This section includes a summary of the unpublished findings in Erasmus School of Law, *Midterm Review 2012. Self-Evaluation Research 2009-2011. Part 1: Management Report*, September 2012.

42. VSNU, *Definitie afspraken Wetenschappelijk Onderzoek*, april 2010.

43. <http://www.law.berkeley.edu/library/iflp/>.

Index<sup>44</sup>, and the Journal Citation Reports<sup>45</sup>. For any new titles to be added to the list it must be demonstrated that the journal (1) features scientific and thus not mainly professional output and (2) features a referee system that is independent from the managing board of the journal. In addition to these faculty-wide general output norms, research programme Directors and frontier research group leaders can establish more demanding output norms for participants, such as with regard to publications in international refereed journals. Failure to meet such standards can result in exclusion from the research programme.

The performance of individual researchers is peer reviewed biennial by an appointed jury for research assessment, consisting of senior researchers with an outstanding publication record, which advises on the granting of the research qualification. As a rule professional publications do not count towards the research qualification. Moreover, research performance forms part of the annual performance appraisals that takes place for all scientific staff. As far as the periodical evaluation of the research activities at institutional level is concerned, the Erasmus School of Law is subject to the six-year evaluation cycle established by the VSNU, as has been mentioned above. Regarding the midterm review, Erasmus School of Law gives preference to the evaluation by an international external evaluation committee rather than a purely internal peer review.

The Erasmus School of Law quality assurance system also extends to PhD research. All PhD candidates employed at the law school participate in the Erasmus Graduate School of Law, which main features are a centralised open selection procedure (based on the quality of the candidates and the relevance of the research proposals for the research profile of the law school), structured scientific training (namely methodology, reflection and practical skills) and regular assessments of the progress by a doctorate committee throughout the PhD trajectory. As part of their education PhD candidates are stimulated to publish articles in refereed international journals. While PhD candidates fall outside the research qualification system described above, different to other researchers, they can be accepted for participation in a research programme or frontier research group even before the defence of their theses.

44. <http://lawlib.wlu.edu/LJ/>.

45. <http://thomsonreuters.com/journal-citation-reports/>.

### III. Strategies towards a sustainable research assessment system for the legal discipline

What has become clear so far is that in the Netherlands, as has been confirmed by the Committee responsible for the 2009 evaluation of legal research, there is no generally shared understanding of scientific quality within the Dutch legal research community, on the basis of which research results can be evaluated<sup>46</sup>. As different definitions, standards and practices are used for the assessment of legal research law schools to a certain extent define their own qualitative and quantitative standards. Put differently, despite the application of standard protocols a generally acknowledged coherent research assessment system that allows for the measurement of quality and/or impact of output is absent.

In fact, the SEP and the definitions offered by the VSNU arguably offer little direction not only with regard to the definition of the output classifications, but more importantly with regard to the actual criteria that should determine the quality of output. Under these circumstances, the evaluation of research is limited to a book-keeping exercise at best. Once a publication has been accounted for as ‘scientific’, like the article presented in paragraph 3, it will keep this qualification in the academic statistics, even though the number of views and downloads in the Repository is low and even if it remains unnoticed in academic literature or legal practice.

A less flattering assessment would be that the evaluation of research based on such imprecise standards is non-transparent and potentially arbitrary, as it relies entirely on the judgments of those in charge of the assessment, be it in the context of internal evaluations at the faculty or University level or external evaluations at individual, i.e. researcher, or institutional level. What is more, comparing the output of individual researchers, research groups or institutions along the lines of what has been done in the context of the 2009 assessment seems highly problematic in such circumstances.

While in the Netherlands the pressure on the legal discipline to introduce a more reliable system to measure output has increased in recent years, there is still no consent on the lines along which a more sustainable research assessment system for legal research could be established. Indeed, should the legal discipline follow other disciplines

46. EVALUATIECOMMISSIE RECHTSWETENSCHAPPELIJKE ONDERZOEK (Commissie Koers), *Kwaliteit en diversiteit* (Report), Amsterdam 2009, p. 54.

in utilizing impact monitoring as a tool to determine excellence, or should the focus come to rest on establishing a system of quality indicators based on peer assessment? The answer to this intricate question does not only depend on what is feasible, but arguably also on what aspect(s) of output should be measured.

*A. Utilizing bibliometric data to measure output in the legal discipline*

At first sight and in particular from the point of view of an outsider, the solution could be to follow the example of those disciplines that utilize impact monitoring as the main tool to evaluate and rank output. Mainly Thomson Reuter's ISI Web of Science and Elsevier's Scopus offer bibliographic and citation databases that do not only provide information on the impact of academic journals, but also on the productivity and impact of individual researchers that publish in such journals. Using this information, it becomes relatively straightforward to assess and compare performance not only at the individual, but also at the institutional level without moreover being limited to any one country or region. Put in a nutshell, it could thus be argued that law schools should embrace the h-index<sup>47</sup> as main measurement instrument to measure the quality and impact of output.

However, a closer look reveals major problems with the application of such a system to the legal discipline due to the scope of the existing big databases and the dominant publication culture in the legal discipline. As to the former, mainstream electronic databases such as ISI Web of Science and SCOPUS focus on English language journals, with a strong bias – certainly in the area of law – towards North American journals. Foreign language (legal) journals are included to a much lesser extent. Yet, a large part of the output in the legal discipline, by the very nature of its research focus, has strong ties with the national domain and is published in non-English outlets, as the audience is largely national. What is more, existing databases focus on journal publications and provide little to no bibliometric data on monographs or edited volumes. These findings are confirmed by a 2011 study by the Dutch Centre for Science and Technology Studies (CWTS) in the area of Humanities and Law, which came to the conclusion that 'a sole dependence on bibliometric techniques as can be applied with ease in the medical, life and

47. J.E. HIRSCH, « An index to quantify an individual's scientific research output », *PNAS* November 15, 2005, Vol. 102 N° 46, 16569-16572.

natural sciences, and with some effort in the engineering sciences, mathematics, statistics and some social sciences (economics, psychology, management science), do not work properly in the social sciences, humanities and law<sup>48</sup>. Underpinning this observation, a 2011 pilot study<sup>49</sup> conducted by the Erasmus University Central Library in close collaboration with the Erasmus School of Law, determined that in the reference period about 90 percentage of the output was not visible in the most widely used rating systems. Not only Dutch legal journals, Dutch or English-language monographs and edited volumes but even a considerable number of international legal journals could not be traced. Similar percentages have also been established for other Dutch law schools. In such circumstances, application of the existing databases to determine the impact would render a large parts of legal output insignificant, and would effectively amount to an unwarranted marginalisation of a whole academic discipline<sup>50</sup>.

From these observations, different and not necessarily concurring conclusions may be drawn. First, true to the motto 'If you can't beat them, join them !', it may be argued that the legal discipline needs to rethink its publication strategy, namely focusing on publications in international journals with an impact factor that are covered by the existing databases. Moreover, more emphasis should then be put on interdisciplinary research, as this gives access to new publication outlets for legal research. Finally, publication in non-English outlets, monographs and edited volumes should be discouraged. At the opposite end of the argument, it may be claimed that for the time being the use of bibliometric instruments is practically impossible considering the large blind spot of existing databases. Moreover, it could be observed that to adjust the publication strategy would mean to discard the specific characteristics of the legal discipline, that distinguish it from other disciplines and that define its large societal relevance namely for policy-makers and practitioners. Traditionally legal scholars in many fields often wait at the practitioner's hand and foot by conducting research with a mainly national or regional and mono-disciplinary legal focus.

48. T. VAN LEEUWEN, E. VAN WIJK and P. WOUTERS, *Bibliometric analysis of output in the Humanities and Law: A case study of Dutch law and history research output 2004-2009*, CWTS, Leiden University; Erasmus School of Law, *Midterm Review* 2012.

49. ERASMUS UNIVERSITY CENTRAL LIBRARY, « Publishing with more impact », *Pilot Study*, 2011.

50. Traditionally, public international law, European (Union) law and comparative law are exceptions in this regard.

Taking the middle road may entail embracing the principle premise that measuring the impact of output should be at the core of a research assessment, while at the same time recognising the shortcomings of the existing databases and the domain-specific publication culture in law. This would then call for the development of an alternative bibliometric toolbox that covers a wider range of output channels, including (non-English) legal journals, as well as monographs and edited volumes. To include the latter two categories makes sense because at least in the legal discipline, monographs and contributions to edited volumes are also a major source of citations of journal articles. A viable strategy could be one that aims at extending existing databases and utilizing alternative sources, including *inter alia*:

- Push for a bigger coverage of law in existing system, e.g. in the contest the European Reference Index for the Humanities (ERIH);
- Mining of existing databases managed by national journal and book publishers;
- Building of national repository in which all output is systematically included;
- Exploring the potential of 'altmetrics'.

Along these lines, for example Erasmus School of Law in collaboration with the Erasmus University Rotterdam Central Library is currently exploring the possibility to build an integrated system combining these alternative sources with the existing databases. At Leiden University a project has been initiated to create a general open access environment, which could be used to assess the quality of the research involved.

To be sure, investing in a tailor-made bibliometric toolbox for law accepts the premise that the focus should be on the ranking of journals, as well as the measuring of quantity of output and citations. Yet, at least in the Dutch academic discourse, serious descending voices have been raised in this regard. Namely relying on the ranking of journals has been met with scepticism. Kortmann for example criticises proponents of such a system for having a strongly distorted and mechanic picture of reality<sup>51</sup>, and Smits notes that in determining who is the best scientist, rankings can only be one aspect<sup>52</sup>.

51. See C.A.J.M. KORTMANN, « Schrijven en Tellen », *Ars Aequi*, maart 2010, pp. 156-157.

52. See J. SMITS, « Ranking van tijdschriften: over de psyche van de wetenschapper en Harvard aan de Rijn », *Ars Aequi*, maart 2010, pp. 164-165.

Indeed, it is difficult to maintain that the ranking of a journal in which a publication appears, the number of publications and the number of citations are adequate to evaluate the quality of research output. In this context, it can be observed that even the fact that a publication is cited by peers cannot in all instances be interpreted as an (implicit) recognition of the (innovative) quality of the publication in question. The latter, it could be argued, requires an assessment of individual research output based on generally accepted quality indicators. As such general research assessment systems can – at best – measure performance, but not individual quality.

*B. Improving the existing system of peer assessment as an alternative?*

What differentiates peer assessment from the use of bibliometric data is the actual review of the scientific quality of individual researchers and research groups that reaches further than the mechanical counting of output and citations. Yet, establishing an objective and comprehensive quality assessment system that allows for a transparent and non-arbitrary assessment of research, is anything but straightforward. Indeed, while the Dutch system is based on peer assessment, the actual criteria that are applied by external reviewers to determine quality are not encoded. The 2009 Dutch legal research assessment highlights how this can result in differences in emphasis in the assessment of individual programmes. There is a call for identifying and defining, and thus objectifying the criteria that are applied in assessing the quality of output. Such criteria *could inter alia* include:

- Methodological justification;
- Research question / hypothesis;
- Originality, significance and rigour, determined with reference to existing relevant research;
- (Outlet);
- (Impact of output, determined by bibliometric data and valorisation indicators)<sup>53</sup>.

To be sure, while such an evaluation may be feasible on an individual basis, e.g. in the context of a performance evaluation or an appointment procedure, it is practically unworkable in the context

53. It could be argued that the last two bullet points are not strictly speaking linked to the assessment of the quality of an individual output.



of a country-wide research assessment considering the amount of publications currently produced in any one given evaluation period.

In the case of the Netherlands, the strategy that has been applied until now to deal with this is twofold. Firstly, the research of individuals has been assigned to research groups, which then become the actual objects of the assessment. Put differently, research is currently assessed on a collective rather than individual basis, even if outstanding individuals may occasionally take centre stage in the assessment. Secondly, assessment has been based on the classification of output<sup>54</sup>. With regard to the latter, the SEP could be improved by providing more detailed definitions and further specifications of subcategories of output classifications. Concerning journals, rather than to generally refer to ‘externally refereed journals’, a term leaving much room for interpretation, a list of Dutch and non-Dutch journals considered to fall into this category should be established. The category ‘books’ is much too broad and unspecified, as it namely does not classify monographs and edited volumes based on the quality of the publishers and whether book proposals are subjected to a blind-referee system. Certainly with regard to international publishers, the top publishers can easily be identified. In the national sphere this is more difficult, if not impossible, due to the small number of and comparability between publishers. Also, a differentiation should be made between A and B journals with regard to editorships.

An alternative strategy could involve a substantial decrease of the amount of output to be externally assessed. Law schools would be limited to submitting x number of key publications per FTE research time, research group or individual researcher for a given reference period. The quality of the research group and research institute would then be assessed based on this sample. The United Kingdom Research Excellence Framework (REF), which relies on a system of ‘discipline-based expert review’ based on research output that is selected and submitted by the research institutions that receive public funding provides an example for such a system<sup>55</sup>. The institutions subject to review themselves thus select what research output is submitted for external evaluation. The three basic assessment criteria applied in the REF are not dissimilar to those applied in the SEP

54. See section II.B.1.

55. REF 2014, Research Excellence Framework, Assessment framework and guidance on submissions, available at <http://www.ref.ac.uk/media/ref/content/pub/assessmentframeworkandguidanceonsubmissions/GOS%20including%20addendum.pdf>, p. 4.

2015-2021, namely *outputs* (defined in terms of the quality of the submitted research output), *impact* (defined in terms of “‘reach and significance’ of impacts on the economy, society and/or culture”) and *environment* (defined as the ‘vitality and sustainability’ of the research environment), which respectively count for 65, 20 and 15 per cent of the overall grade given to a specific research output<sup>56</sup>.

#### IV. Conclusions

What has been highlighted in this contribution is that despite the existence of a standard evaluation protocol for the assessment of research, there is no generally shared understanding of scientific quality within the Dutch legal research community, on the basis of which research results can be evaluated. It is thus little surprising that law schools to a certain degree define their own qualitative and quantitative standards, thereby filling the lacunae. What is more, a generally accepted method to measure quality is missing. Namely there is a notable absence of generally shared quality indicators, such as the journal rankings and citation indexes that are used in other disciplines. At the same time, the evaluation of the intrinsic quality of a publication can hardly be determined based only on broad classifications of the publication form, medium (journals, book etc.) and/or the language in which a publication is written.

Notwithstanding these shortcomings, no serious attempts have been made until now to follow the example of other (social sciences) disciplines to monitor impact as the main tool to evaluate and rank output. This may be little surprising, given that utilizing existing databases to monitor impact would have rather disastrous consequences for the legal discipline, due to the current scope of such databases and the predominant publication culture in the legal discipline.

As Gutwirth has put it: ‘To negatively evaluate a [...] legal researcher because he has too few ISI publications is the same as saying that a Thai restaurant is no good because there are no chips or pizzas on the menu’<sup>57</sup>.

56. *Ibid.*, p. 6.

57. Brackets added. S. GUTWIRTH, « The evaluation of legal science. The V.I.R.-model for integral quality assessment of research in law: what next? », in H. EISENDRATH and J.P. VAN BENDEGEM (eds.), *It takes two to do science. The puzzling interactions between science and society* (Brussels, VUBPRESS, 2009), 69-80.

However, from this it should not be concluded that impact monitoring can under no circumstances ever be applied to the legal discipline. What has been suggested in this contribution is that a double strategy involving influencing publication behaviour and the building of a discipline-specific database may be the way forward. At the same time, peer assessment can fulfil an important role in ensuring that the evaluation of quality does not deteriorate to the mechanical counting of output and citations.

Presently the draft paper to be found at [http://works.bepress.com/cgi/viewcontent.cgi?article=1015&context=serge\\_gutwirth](http://works.bepress.com/cgi/viewcontent.cgi?article=1015&context=serge_gutwirth), para. 13, is cited.

